

1 **Insufficiency of PCR diagnostics for Detection of Diarrhoeagenic *Escherichia coli* in Ibadan,**  
2 **Nigeria**

3

4 Olabisi C. Akinlabi<sup>1</sup>, Rotimi A. Dada<sup>2</sup>, El-shama Q. Nwoko<sup>1</sup>, and Iruka N. Okeke<sup>1</sup>.

5

6 1. Department of Pharmaceutical Microbiology, Faculty of Pharmacy, University of Ibadan, Oyo,  
7 Nigeria.

8 2. Medical Laboratory Science Programme, College of Health Sciences, Bowen University Iwo  
9 and Department of Pharmaceutical Microbiology, Faculty of Pharmacy, Ahmadu Bello University,  
10 Zaria.

11

12 **\*Correspondence to** [iruka.n.okeke@gmail.com](mailto:iruka.n.okeke@gmail.com).

13

## 14 **Abstract**

15 Understanding the contribution of different diarrhoeagenic *Escherichia coli* pathotypes to disease  
16 burden is critical to mapping risk and informing vaccine development. Targeting select virulence  
17 genes by PCR is the diagnostic approach of choice in high-burden, least-resourced African  
18 settings. We compared the performance of a commonly-used multiplex protocol to whole genome  
19 sequencing (WGS).

20 PCR was applied to 3,815 *E. coli* isolates from 120 children with diarrhoea and 357 healthy  
21 controls. Three or more isolates per specimen were also Illumina-sequenced. Following quality  
22 assurance, ARIBA and Virulencefinder database were used to identify virulence targets. Root  
23 cause analysis of deviant PCR results was performed by examining target sensitivity using  
24 BLAST, Sanger sequencing false-positive amplicons, and identifying lineages prone to false-  
25 positivity using in-silico multilocus sequence typing and a Single Nucleotide Polymorphism  
26 phylogeny constructed using IQTree.

27 The sensitivity and positive predictive value of PCR compared to WGS ranged from 0-77.8%  
28 while specificity ranged from 74.5-94.7% for different pathotypes. WGS identified more  
29 enteroaggregative *E. coli* (EAEC), fewer enterotoxigenic *E. coli* (ETEC) and none of the Shiga  
30 toxin-producing *E. coli* detected by PCR, painting a considerably different epidemiological  
31 picture. Use of the CVD432 target resulted in EAEC under-detection, and enteropathogenic *E. coli*  
32 *eae* primers mismatched more recently described intimin alleles common in our setting. False  
33 positive ETEC were over-represented among West Africa-predominant ST8746 complex strains.

34 PCR precision varies with pathogen genome so primers optimized for use in one part of the world  
35 may have noticeably lower sensitivity and specificity in settings where different pathogen lineages  
36 predominate.

37

38

## 39 Introduction

40 Diarrhoea is a principal contributor to morbidity and mortality in young children, particularly in  
41 Africa and parts of Asia. A range of pathogens can produce diarrhoea, including diarrhoeagenic  
42 *Escherichia coli* (DEC). *E. coli* is the most common proteobacterial intestinal commensal and this  
43 presents a challenge for delineating potentially pathogenic strains, which harbor virulence genes,  
44 from those that are beneficial. There are several DEC pathotypes, five of which have been  
45 validated in human volunteer challenge studies and/or significant outbreaks: enterotoxigenic *E.*  
46 *coli* (ETEC), enteropathogenic *E. coli* (EPEC), Shiga toxin-producing *E. coli* (STEC) (which  
47 includes enterohaemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC) and enteroaggregative  
48 *E. coli* (EAEC) (1). Diffusely-adherent *E. coli* (DAEC), cell-detaching *E. coli* (CDEC) and  
49 cytolethal distending toxin-producing *E. coli* (CLDTEC) are among pathotypes that have been  
50 proposed but require more evidence to validate them. Within each pathotype, there is considerable  
51 genetic diversity and different lineages evolved convergently. There are also hybrid strains that  
52 carry virulence genes corresponding to multiple pathotypes (2–7).

53

54 Traditionally, DEC pathotype identification uses tissue-culture or vertebrate animal-based  
55 diagnostics, which are onerous, highly specialized and therefore unsuitable for routine clinical use  
56 (8). The Gold Standard for adherent DEC pathotypes (EPEC, EAEC, DAEC and CDEC) has been  
57 a tedious epithelial adherence assay, which few labs can perform, while ETEC, STEC and  
58 CLDTEC Gold Standards depend on detection of the toxins that these pathotypes produce. EIEC  
59 can be identified by detecting entry into epithelial cells or conjunctivitis produced in rodent  
60 models. Today, the genes encoding defining phenotypes for each pathotype are known, making it  
61 possible to detect DEC using DNA-based methods that offer greater throughput and more

62 widespread application. PCR for DEC in particular is a low-cost method that can leverage  
63 instrumentation and expertise initially acquired for other purposes and therefore present in most  
64 reference and research laboratories, and in some smaller clinical laboratories. While automated  
65 PCR-based protocols are rapidly gaining ground in high-income settings, the low cost for manual  
66 PCR set up, particularly for end-point PCR, and the flexibility around its consumable procurement  
67 mean that it remains an approach of choice in low-income settings, including Nigeria where we  
68 work. Because of the large number of targets needed to cover the most prominent DEC pathotypes,  
69 and the need to integrate DEC diagnosis into a larger search for diarrhoeal pathogens, multiplex  
70 PCR protocols are favored, even though these are notoriously challenging to optimize (9–11). PCR  
71 is also perceived to have a lower false-positive rate than nucleotide hybridization, which was  
72 popular in the twentieth century but requires somewhat costly and often inaccessible detection  
73 methods. Moreover, hybridization can pick up similar but not identical genes and therefore  
74 confound DEC diagnoses (12,13)

75  
76 The range of known virulence gene alleles and pathogenic lineages, as well as the absence of a  
77 precise genetic definition for EAEC and newer pathotypes, mean that the more targets that can be  
78 detected, the more sensitive a testing method becomes. For all of these reasons, whole genome  
79 sequencing offers the promise of reliably detection any DEC pathotype, without the need of several  
80 cumbersome assays. Due to the high cost and limited availability of WGS, most labs continue to  
81 use PCR-based tests that were developed and validated in a very limited number of settings against  
82 small panels of strains. We compared the performance of a popular PCR-based protocol that has  
83 been in use for almost two decades to WGS. We then identified root causes of PCR under-

84 performance in our setting thereby uncovering more general principles for PCR diagnostics failure  
85 for pathogens that show considerable diversity.

86

## 87 **Materials and methods**

88 **Ethical considerations:** This study was part of a larger investigation of the epidemiology of  
89 diarrhoeagenic *E. coli* among children with diarrhoea presenting to primary health clinics in  
90 southwestern Nigeria with ethical approval from the University of Ibadan/University College  
91 Hospital ethics committee (UI/EC/15/0093). Patient parents or guardians provided written  
92 informed consent to participate. According to the approved protocol, neither PCR nor WGS  
93 results, which were performed for surveillance and not obtained in a timely manner, were used to  
94 inform patient care.

95

96 **Review of multiplex PCR methods:** We systematically searched Pubmed  
97 <https://pubmed.ncbi.nlm.nih.gov> for “diarrhoeagenic *Escherichia coli*” and “PCR method”.  
98 Abstracts retrieved were read to determine whether they included development of a protocol, and  
99 if so, where the protocol was developed. Where this was the case, the number of papers citing the  
100 reference in question and the location of cited studies was retrieved using Google Scholar  
101 <https://scholar.google.com>.

102

103 **Strains:** Strains used in this study were collected as part of an epidemiological survey for  
104 diarrhoeagenic *E. coli* (6). Up to ten isolates were obtained per stool specimen and all isolates  
105 were subjected to PCR screening. Up to three isolates showing different biochemical and/or

106 antimicrobial susceptibility profiles were whole genome-sequenced. Control strains, which are  
107 well-studied and genome-sequenced DEC from other parts of the world, are listed in Table 1.

108

**Table 1: PCR oligonucleotides from the Aranda et al protocol that were used in this study**

109

| Primer Designation | Function encoded by target gene | Primers (5' to 3')     | Target gene or probe | Amplicon size (bp) | Positive control strains |
|--------------------|---------------------------------|------------------------|----------------------|--------------------|--------------------------|
| <i>eae f</i>       | Intimin                         | CTGAACGGCGATTACGCGAA   | <i>eae</i>           | 917                | E2348/69                 |
| <i>eae r</i>       | EHEC and EPEC                   | CCAGACGATACGATCCAG     |                      |                    |                          |
| <i>bfp f</i>       | Bundle-forming pilus of EPEC    | AATGGTGCTTGCGCTTGCTGC  | <i>bfpA</i>          | 326                | E2348/69                 |
| <i>bfp r</i>       |                                 | GCCGCTTTATCCAACCTGGTA  |                      |                    |                          |
| <i>CVD432 f</i>    | Dispersin transporter           | CTGGCGAAAGACTGTATCAT   | <i>aatA</i> (CVD432) | 630                | 042,17-2,60A             |
| <i>CVD432 r</i>    |                                 | CAATGTATAGAAATCCGCTGTT |                      |                    |                          |
| <i>LT f</i>        | Heat labile toxins of ETEC      | GGCGACAGATTATACCGTGC   | <i>elt</i> (LT)      | 450                | H10407                   |
| <i>LT r</i>        |                                 | CGGTCTCTATATCCCTGTT    |                      |                    |                          |
| <i>ST f</i>        | Heat stable toxins of ETEC      | ATTTTTMTTCTGTATTRTCTT  | <i>est</i> (ST)      | 190                | H10407                   |
| <i>ST r</i>        |                                 | CACCCGGTACARGCAGGATT   |                      |                    |                          |



---

|               |                  |                            |             |     |        |
|---------------|------------------|----------------------------|-------------|-----|--------|
| <i>ipaH f</i> | invasion-        | G TTCCTTGACCGCCTTTCCGATACC | <i>ipaH</i> | 600 | E137   |
| <i>ipaH r</i> | associated locus | GTC                        |             |     |        |
|               | of EIEC and      | GCCGGTCAGCCACCCTCTGAGAGTA  |             |     |        |
|               | <i>Shigella</i>  | C                          |             |     |        |
| <i>stx1 f</i> | Shiga toxin 1    | ATAAATCGCCATTCGTTGACTAC    | <i>stx1</i> | 244 | EDL933 |
| <i>stx1 r</i> | and 2 of EHEC    | AGAACGCCCACTGAGATCATC      |             |     |        |
|               | and STEC         |                            |             |     |        |
| <i>stx2 f</i> | Shiga toxin 1    | GGCATGTCTGAAACTGCTCC       | <i>stx2</i> | 190 | EDL933 |
| <i>stx2 r</i> | and 2 of EHEC    | TCGCCAGTTATCTGACATTCTG     |             |     |        |
|               | and STEC         |                            |             |     |        |

---

111  
112 **Virulence Gene Detection by PCR:** Molecular diagnosis of DEC was performed by gene  
113 amplification using PCR. The *E. coli* genomic DNA was screened by multiplex PCR by targeting  
114 the genes as described by Aranda *et al.*, 2004 (9) (Table 1). However, the protocol was modified  
115 due to the closeness of the band sizes in Aranda's PCR2 targets of *est* and *stx2* genes with band  
116 sizes 180 and 190 bp respectively (14). The specific modification splits PCR2 into two separate  
117 reactions, designated PCR3 and PCR 4 (Table 2). Likewise, because of previously observed non-  
118 specific products, the annealing temperature of the PCR reactions was increased and the annealing  
119 time was reduced for each reaction to avoid previously detected non-specific products (14). The  
120 conditions used are as outlined in Table 2. All multiplex PCR was performed using the  
121 illustra™ PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare UK.). These  
122 contain *Taq* polymerase, BSA, stabilizers, dNTPs, 1,5nM MgCl<sub>2</sub>, 50mM KCl and reaction buffer.  
123 Each beads is reconstituted to 25 µL by adding 2.5 µl of each forward and reverse primer solution,  
124 20 ng/µl of *E. coli* DNA and nuclease-free water to complete the final volume. Well known  
125 prototypical strains were used as positive controls (Table 1) and DH5α and CFT073 as negative  
126 controls in the PCR reactions.

127  
128 Amplicons were visualized on TAE agarose gels containing 1% agarose for PCR 1 and PCR 3 and  
129 1.5% for PCR 4. The volume of agarose used provided 4 mm depth in the electrophoresis tray.  
130 Gels were visualized and photographed using the UVP® GelMax Imager which utilizes a built-in  
131 midrange 302 nm ultraviolet (UV) transilluminator.

132

133 **Table 2: Polymerase Chain Reaction Cycle**

| Step                 | PCR 1                                       | PCR 2  | PCR 3   | PCR 4   |
|----------------------|---|--|---|---|
|                      | Aranda <i>et al.</i> ,<br>2004 (9)          | Aranda <i>et al.</i> ,<br>2004 (9)                                     | Aranda <i>et al.</i> ,<br>2004 (9) ,<br>Odetoyin <i>et al.</i> ,<br>2016 (14) | Aranda <i>et al.</i> ,<br>2004 (9)<br>Odetoyin <i>et al.</i> ,<br>2016 (14) |
| Target genes         | <i>aatD</i> , <i>eae</i> and<br><i>bfpA</i> | <i>elt</i> , <i>est</i> , <i>ipaH</i> ,<br><i>stx1</i> and <i>stx2</i> | <i>elt</i> , <i>est</i> and<br><i>ipaH</i>                                    | <i>stx1</i> and <i>stx2</i>   |
| 1. Hot start         | 94°C 3min                                   | 94°C 3min  | 94°C 3min   | 94°C 3min   |
| 2. Denaturing        | 94°C 30sec                                  | 94°C 45sec   | 94°C 45sec  | 94°C 45sec  |
| 3. Annealing         | 60°C 30 sec                                 | 51°C 30 sec  | 51°C 30 sec   | 58°C 30 sec   |
| 4 Extension          | 72°C 1 min                                  | 72°C 45 sec  | 72°C 45 sec   | 72°C 45 sec   |
| 5. Cycling           | Steps 2-4<br>repeated 40<br>Times           | Steps 2-4<br>repeated 40<br>times                                      | Steps 2-4<br>repeated 40<br>times   | Steps 2-4<br>repeated 40<br>times   |
| 6 Terminal extension | 72°C 7 min                                  | 72°C 7 min   | 72°C 7 min  | 72°C 7 min  |

135

136

137

138 **Whole genome Sequencing:** *E. coli* genomic DNA were extracted using the Wizard Genomic  
139 Extraction kit (Promega) according to the manufacturer's instructions and they were shipped and  
140 bulk sequenced in Sanger using the illumina platform as described previously (6). Raw data were  
141 assembled by SPAdes assembler and assembly quality was determined using Quality Assessment  
142 Tool for Genome Assemblies (QUAST), and CheckM. ARIBA and Virulencefinder database,  
143 were used to identify virulence genes. Intimin allele calling was performed by comparing database  
144 prototypes to assembled genomes using BLAST. Sequence data were submitted to ENA and are  
145 available from ENA <https://www.ebi.ac.uk/ena/browser/home> and  
146 Genbank <https://www.ncbi.nlm.nih.gov/genbank/> as Bioproject PRJEB8667.

147  
148 **PCR amplicon sequencing:** Selected isolates that amplified the virulence genes by PCR but did  
149 not harbor such genes by WGS were reamplified and amplicons were Sanger sequenced at Inqaba  
150 Biotec (Ibadan, Nigeria and Pretoria, South Africa). Returned sequences yielding single peaks  
151 were compared to the NCBI nucleotide database using BLAST to retrieve the match and flanking  
152 sequences. Primer sequences were then compared for within amplicon/ flanking sequence matches.

153  
154 **Mapping PCR diagnostic results onto phylogenies.**  
155 All sequenced isolate reads were mapped to EAEC\_042 reference genome (accession number:  
156 FN554766) to create a multi-fasta alignment file. Using the multi-fasta alignment file, Single  
157 Nucleotide Polymorphism (SNP) were called using SNP-sites and the SNP file was used to  
158 construct a phylogenetic tree using IQTree software (15). Microreact (16) was used to visualize  
159 phylogenetic trees, STs and loci identified by PCR and WGS.

160

161 **Statistical Analysis:** Epi Info version 7 software (Centers for Disease Control and Prevention,  
162 Atlanta, GA, USA) was used for statistical calculations. Fisher’s exact test and Chi square test  
163 were used to test the association of test-defined pathotypes with disease. p-values < 0.05 were  
164 considered significant, with Bonferroni corrections used where applicable

165

## 166 **Results**

### 167 **Multiplex PCR for DEC is rarely developed in Africa but commonly used to identify DEC in**

168 **African settings:** A search for “diarrheogenic *Escherichia coli*” and “PCR method” on Medline

169 (Pubmed) initially yielded 207 hits, which when reviewed uncovered 16 studies that developed

170 multiplex PCR protocols for diarrhoeagenic *E. coli* (Table 1). In contrast to some other diagnostic

171 protocols, where one or a few methods gain prominence over time, there are a wide variety of

172 published PCR protocols for DEC with 7 of those published before 2010 having 1,132 or more

173 citations (Table 3). A similar number of studies developed after 2010 are accumulating citations

174 as well. Five protocols were developed and published from Europe, four protocols were developed

175 by groups working in Asian countries and three by South American groups. Only two of the

176 protocols were published by groups based in Africa (both in Egypt) and none in Nigeria, West

177 Africa or elsewhere in the WHO Africa region (Table 3). While target choice has been compared

178 across different protocols (17), we were not able to find studies that included a validation on

179 isolates from the WHO-AFRO region nor any studies comparing different PCR protocols with one

180 another.

181

182 Table 3: Application of published DEC PCR identification protocols on the African continent

| Protocol (Reference)     | Developed in | #Google Scholar citations as at 1 Dec 2022 | # Citations from African use study | Location of African studies citing the protocol*  |
|--------------------------|--------------|--|------------------------------------|---|
| Toma et al 2003(10)      | Japan        | 297  | 11                                 | Cote d'Ivoire, Kenya, Nigeria, Tanzania, South Africa, Uganda                             |
| Aranda et al 2004(9)     | Brazil       | 294  | 18                                 | Ghana, Kenya, Nigeria, South Africa   |
| Vidal et al 2005(18)     | Chile        | 301  | 24                                 | Bukina Faso, Chad, Ethiopia, Egypt, Ghana, Kenya, Nigeria, South Africa and multi-country |
| Muller et al 2007(19)    | Germany      | 56   | 0                                  |   |
| Metwally et al 2007(20)  | Egypt        | 22   | 2                                  | Egypt, Nigeria  |
| Antikainen et al 2009(8) | Finland      | 93   | 15                                 | Bukina Faso, Chad, Malawi, South Africa   |
| Tobias et al 2012(17)    | Sweden       | 69   | 2                                  | Ghana, South Africa   |
| Hegde et al 2012(21)     | India        | 70   | 4                                  | Burkina Faso, Ethiopia, Nigeria, Sudan  |
| Fialho et al 2013(22)    | Brazil       | 12   | 0                                  |   |

|                           |          |    |    |  |
|---------------------------|----------|----|----|--|
| Helmy et al<br>2013(23)   | Egypt    | 7  | 1  | Nigeria  |
| Al Talib 2014(24)         | Malaysia | 18 | 0  |  |
| Oh et al 2014(25)         | Korea    | 22 | 0  |  |
| Onori et al 2014(26)      | Italy    | 80 | 3  | Ghana, South Africa, Sudan                                 |
| Naziri et al 2015(27)     | Iran     | 0  | 0  |  |
| Sjoling et al<br>2015(28) | Sweden   | 37 | 15 | Burkina Faso, Cote D'Ivoire,<br>Ghana, Kenya, South Africa |
| Zhang et al 2020(29)      | China    | 15 | 0  |  |

184

185

186 Within Africa, we found clear geographic bias in protocol selection that could not, from the  
187 aforementioned, be attributed to validation, among studies seeking DEC in human, food or  
188 environmental specimens (Table 3). Some well-cited protocols (17,19,26,29) have rarely or never  
189 been used in Africa. The Vidal et al 2005 (18) protocol has been used in the most African countries  
190 but not more than two or three times in any single country. This study evaluated the Aranda et al  
191 (2004) (9) protocol, which uses a minimal primer set to identify all five pathotypes that have been  
192 validated through human volunteer studies and/or outbreaks. This was the most commonly  
193 employed protocol in Nigeria and some of its primer pairs are employed in other popular multiplex  
194 protocols (17,22,30). The protocol originally presented by Aranda et al (2004) (9) consists of two  
195 multiplex PCRs that detected a total of eight targets, all validated DEC pathotypes. Noting that  
196 the amplicon sizes for *stx1* (180 bp) *elt* (190 bp), we previously separated PCR2 into two reactions  
197 to avoid errors of interpretation (31).

198

199 **The epidemiological picture for the same specimen set is radically different when DEC are**  
200 **identified by PCR and WGS:** The Aranda et al (2004)(9) PCR performed well for all positive  
201 and negative control strains, which were included in every batch, reproducibly giving the expected  
202 result every time. *E. coli* from 413 specimens obtained from children in Ibadan, totaling 3,815  
203 isolates and averaging eight isolates per specimen, were tested by PCR. All PCR positives were  
204 sequenced and in instances where there were <2 such positives from any specimen, three isolates  
205 with different biochemical profiles were selected for sequencing. Thus, due to resource limitation,  
206 a smaller number of 1,230 (averaging three per specimen) were tested by whole genome  
207 sequencing but all pathotypes identified by PCR were sequenced for validation. As shown in Table



208 4, in spite of this overlap, the epidemiological picture revealed by PCR is very different from that  
209 defined by WGS.

210  
211 Table 4 shows that PCR detected all five pathotypes and identified ETEC 150 (31.4%) as the most  
212 common pathotype (31 *est*-positive strains, 206 *elt*-positive strains and 67 strains having both  
213 targets), followed by CVD432-positive EAEC 99(20.8%), STEC positive *stx1* and/or *stx2*  
214 46(9.6%), *eae* and/or *bfp*-positive EPEC 43(9%) and EIEC from which *ipaH* amplified 33(6.9%).  
215 Thus PCR detected all five DEC pathotypes as shown in Figure 1 and ETEC with *est* and *elt* genes  
216 or *elt* alone were significantly associated with diarrhoea ( $p < 0.05$ ), based on PCR results. The *stx1*  
217 and *stx2* genes were detected in 12 and 63 isolates, respectively, by PCR. WGS did not verify any  
218 of the *E. coli* strains as harboring a Shiga toxin gene. Thus, PCR identified 70 isolates that could  
219 be classified as STEC but none of the PCR-identified ‘STEC’ were verified by whole genome  
220 sequencing. No EHEC (*eae*-positive STEC) were identified by either method.

221  
222 We compared sensitivity and specificity of PCR to WGS for each of the pathotypes. The CVD432  
223 (*aatA*) gene was targeted for EAEC, *eae* and *bfp* for EPEC, *elt* and *est* for ETEC, *ipaH* for EIEC,  
224 *stx1* and *stx2* for STEC, Table 5 shows that PCR sensitivity for DEC identification was 77.8% for  
225 ETEC, 61.7% for EAEC, 45% for EPEC and 0% for both EIEC and STEC. PCR did show better  
226 specificity for EAEC (91.6%), EPEC (94.7%), EIEC (93.7%), STEC (90.1%). ETEC specificity  
227 was considerably lower (74.5%) (Table 5)

228

229

230 **Table 4: DEC pathotypes identified by WGS and PCR from the case and control specimens**

| Pathotypes                       | Cases (%)<br>n=120 |                 | Controls (%)<br>(n=357) |                  | Total specimens<br>(n=477) |            | P-values      |               |
|----------------------------------|--------------------|-----------------|-------------------------|------------------|----------------------------|------------|---------------|---------------|
|                                  | PCR                | WGS             | PCR                     | WGS              | PCR                        | WGS        | PCR           | WGS           |
| <b>EAEC (CVD432)</b>             | <b>25(20.8)</b>    | <b>35(29.2)</b> | <b>74(20.7)</b>         | <b>102(28.6)</b> | <b>99</b>                  | <b>137</b> | 0.9804        | <b>0.9075</b> |
| <b>EHEC/STEC</b>                 | <b>9(7.5)</b>      | <b>0(0)</b>     | <b>39(10.9)</b>         | <b>0(0)</b>      | <b>46</b>                  | <b>0</b>   | 0.2807        | <b>0</b>      |
| <i>stx2</i>                      | 8(6.7)             | 0(0)            | 35(9.8)                 | 0(0)             | 43                         | 0          | 0.2992        | 0             |
| <i>stx1</i>                      | 1(0.8)             | 0(0)            | 1(0.2)                  | 0(0)             | 2                          | 0          | 0.4176        | 0             |
| <i>stx1, stx2</i>                | 1(0.8)             | 0(0)            | 3(0.8)                  | 0(0)             | 4                          | 0          | 1             | 0             |
| <b>EPEC</b>                      | <b>15(12.5)</b>    | <b>7(5.8)</b>   | <b>28(8.4)</b>          | <b>16(4.5)</b>   | <b>43</b>                  | <b>23</b>  | 0.1233        | 1.3202        |
| <i>eae+ bfp+</i>                 | 1(0.8)             | 1(0.8)          | 3(0.8)                  | 2(0.6)           | 4                          | 3          | 1             | 1             |
| <i>eae</i>                       | 1(0.8)             | <b>7(5.8)</b>   | 12(3.4)                 | <b>16(4.5)</b>   | 13                         | 23         | 0.2001        | 0.5499        |
| <i>bfp</i>                       | 6(5)               | 0(0)            | 13(3.6)                 | 0(0)             | 19                         | 0          | 0.5103        | 0             |
| <b>ETEC</b>                      | <b>41(34.2)</b>    | <b>8(6.7)</b>   | <b>109(30.5)</b>        | <b>11(3.1)</b>   | <b>150</b>                 | <b>19</b>  | 0.4582        | <b>0.0823</b> |
| <i>elt+est+</i>                  | 12(10)             | 1(0.8)          | 15(4.2)                 | 0(0)             | 27                         | 1          | <b>0.0174</b> | 0.2516        |
| <i>elt</i>                       | 26(21.7)           | 7(5.8)          | 49(13.7)                | 11(3.1)          | 75                         | 8          | <b>0.0387</b> | 0.1711        |
| <i>est</i>                       | 5(4.2)             | 2(1.7)          | 10(2.8)                 | 0(0)             | 15                         | 2          | 0.5446        | 0.0629        |
| EIEC/Shigella<br>( <i>ipaH</i> ) | 9                  | 0(0)            | 22(6.2)                 | 1(0.3)           | 31                         | 1          | 0.6071        | 0             |

231

232 **Table 5: Evaluation of PCR performance for identification of diarrhoeagenic *E. coli***  
233 **compared with sequence data.**

234

|                               | <b>EAEC</b> | <b>ETEC</b> | <b>EPEC</b> | <b>EIEC</b> | <b>STEC</b> |
|-------------------------------|-------------|-------------|-------------|-------------|-------------|
| % Sensitivity                 | 61.7        | 77.8        | 45.0        | 0.0         | ND*         |
| % Specificity                 | 91.6        | 74.5        | 94.7        | 93.7        | 90.1        |
| Positive predictive value (%) | 73.9        | 10.7        | 27.3        | 0.0         | ND*         |
| Negative predictive value (%) | 86.1        | 98.8        | 97.5        | 99.8        | 100         |

235

236 \*There were no STEC true positives

237

238

239

240 **Low sensitivity of EAEC detection by PCR is due to too few targets:** The Aranda et al (2004)  
241 protocol, like most other DEC multiplex protocols, seeks only one EAEC target, CVD432. Whole  
242 genome sequencing on the other hand allows the investigator to screen strains for a total of 14 loci  
243 associated with EAEC (6,32), allowing for greater sensitivity and accounting for the extremely  
244 low sensitivity of PCR. When we compared the prevalence of CVD432 detected by PCR to that  
245 from WGS among isolates that were tested for both. PCR identified 260 (21%) EAEC isolates  
246 while WGS identified 355(28.7%).

247  
248 **Reduced specificity for *bfpA* PCR and reduced sensitivity of *eae* primers:** A prominent red  
249 flag in the PCR results was the 32 strains that repeatedly tested positive for *bfpA* but were *eae*  
250 negative (Figure 1). Typical EPEC strains harbor the chromosomal locus of enterocyte effacement  
251 (LEE) marked by *eae* and the EPEC adherence factor (EAF) plasmid-borne *bfp* genes whilst  
252 atypical EPEC have *eae* on the LEE without *bfp* and the EAF plasmid. Strains carrying *bfp* without  
253 *eae* are rarely reported. In this study, all 40 isolates from which *bfpA* was amplified did not have  
254 the gene as determined by WGS. This includes seven *eae*-positive strains that were misclassified  
255 as EPEC by PCR. A total of 19 atypical EPEC strains (*eae*-positive and *bfpA*-negative) were not  
256 identified by PCR. WGS and screening by VirulenceFinder revealed that 17 of these strains carried  
257 all other LEE markers. We therefore hypothesized that they may carry intimin alleles that are not  
258 amplified by the *eae* primer pair. We used the *eae* gene from prototypical EPEC strain E2348/69  
259 to retrieve *eae* from each of the assembled genomes of these strains using BLAST. BLAST was  
260 then used to type each intimin allele against database prototypes (Table 6). We additionally  
261 examined each allele for complementarity to the *eae1* and *eae2* primers. The results of these  
262 analyses are summarized in Table 6 and show that while seven of the strains sequenced carried

263 intimin alpha, kappa and lambda alleles with no mismatches in the primer annealing sites, and  
264 therefore should have amplified, ten strains carried the iota, rho and theta intimin alleles, which  
265 have between one and three mismatches with the primers, which have the possibility of reducing  
266 amplification at conditions optimal for other alleles. Of note is the fact that in addition to the  
267 Aranda et al (9) protocol, these *eae* primers also feature in multiplex PCRs for DEC by Nazari et  
268 al, Puno- Sarmiento et al, and Dias et al (33–35) and so this problem is likely to occur with some  
269 other multiplex protocols.

270

271 Table 6: Intimin alleles of LEE-positive EPEC strains not detected by PCR.

272

| Strain  | Intimin allele | Prototype accession number | Sequence of <i>eae1</i> and <i>eae2</i> priming sites (mis-matched bases are not in bold) |
|---|----------------|----------------------------|---|
| E2348/69 (control)                                    | alpha (28)     | AF022236.1                 | CTGAACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG  |
| LKD064A   | rho            | DQ523613.1                 | CTGAACGGCGATTACGCAA...880 nt...CTGGATCGTATCGTCTGG   |
| LLH22C  | alpha          | AF022236.1                 | CTGAACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG  |
| LLH35B  | alpha          | AF022236.1                 | CTGAACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG  |
| LLH012D   | kappa          | KT591304.1                 | CTGAACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG  |
| LLD100C   | iota2          | AB647371.1                 | CTGAACGGCGATTACGCAA...880 nt...CTGGATCGTATCGTCTGG   |
| LLH332A, LLH332B, LLH332E                             | episilon       | AB647573.1                 | CTGAACGGCGATTACGCGA...880 nt...CTGGATCGTATCGTCTGG   |
| LL197, LLH161I, LLH161F                               | lambda         | AJ715409.1                 | CTGAACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG  |
| LLH336B, LLH336E, LLD104C, MNH109F, LLHO33E, LLH261C1 | theta2         | FM872418.1                 | TGCACGGCGATTACGCGAA...880 nt...CTGGATCGTATCGTCTGG   |

273

274 **False positivity with *elt* primers is due to non-specific amplification in unfamiliar lineages:**

275 The ETEC pathotype is well-defined with highly conserved virulence targets, the *elt* and *est* genes.  
276 It is also a vaccine priority and there is therefore considerable interest in understanding ETEC  
277 epidemiology in lesser surveilled parts of the world, including Nigeria. In this study, false positives  
278 for most genes were scattered across the phylogeny (Figure 2). This was true to some extent for  
279 *elt*. However, more *elt* false-positivity was seen within phylogroup D and in particular in the  
280 ST720 and ST69 complexes. ST69 is a well-characterized ST complex that includes  
281 extraintestinal and enteroaggregative *E. coli* strains. False-positive ETEC were detected in ST 69  
282 (10/31 isolates) and ST394 (4/12 isolates). The false positivity rate of 14/50 was significantly  
283 greater than the overall false positivity rate ( $p=0.03$ ). ST720 complex is less well studied and in  
284 this study comprised isolates belonging to ST 8746 ( $n=51$ ), ST 720 ( $n=13$ ), ST8749 ( $n=8$ ), ST8131  
285 and ST1136 (4 each), with the highest number of isolates and greatest false-positive rates in  
286 ST8646 (17/51) and ST720 (6/13). Altogether 17/51 ST8746 strains gave a false-positive *elt*  
287 result, as compared to 213 (92.6%) overall ( $p= 0.0006$ ) and the false-positivity rate in ST720  
288 (6/13) was even greater. Of the 17 ST8746 complex strains that were erroneously identified as  
289 ETEC by PCR, all but three were in fact EAEC strains (Figure 2). Thus, the PCR method for this  
290 same lineage simultaneously overestimated the prevalence of ETEC, underestimated EAEC and  
291 hid from existence a previously unreported EAEC lineages. STs are numbered chronologically  
292 with STs in the Wirth et al (2006) (36) Achtman *E. coli* scheme above 1500 recorded after the year  
293 2010. Thus, most STs within the complex, including ST8746, are relatively new to science. Of  
294 interest is that ST8746 itself was the fourth commonest ST we detected in this study, showing that  
295 it is very common in our setting. One strain in the complex, an ST720 strain from Germany, had  
296 been recorded in the database as at 2010. The relatively recent registration of other isolates in the

297 ST complex is likely due to the very few sequence-based studies that take place in Nigeria and  
298 nearby locales. Other ST8746 strains have been reported from Ghana (n=1 ) and United Kingdom  
299 (n= 1 ) (37,38) but the largest existing collection is from our own study. The data point to the  
300 likelihood that false positives with the *elt* primers will be common in Nigeria and other settings  
301 where ST8746 organisms predominate but likely less so in areas where this lineage is absent or  
302 less prominent. They also illustrate the weakness of PCR methods primarily used in West Africa  
303 that are not developed here.

304

### 305 **The *stx* multiplex produces a confounding non-specific amplicon due to mis-priming**

306 In previous studies, we have de-complexed Aranda et al (2004)'s second multiplex, which was  
307 originally designed to contain five primer pairs (14) Thus, we use one multiplex to detect the  
308 *elt*, *est* and *ipaH* genes for ETEC and EIEC and a second duplex PCR for *stx1* and *stx2*. The *stx*  
309 duplex PCR reaction identified 12 *stx1*- and 63 *stx2*-positive strains. Upon WGS analysis, none  
310 of these isolates, nor any of the other isolates sequenced in the study carried these genes. We  
311 sought to understand why PCR yielded a right-sized product in the *stx* duplex and attempted  
312 Sanger sequencing of a few amplicons, most of which failed due to multiple peaks suggesting  
313 that each band consisted of multiple amplicons. We did get a product for the *stx* duplex PCR4  
314 that was presumed to be the right size for *stx1* from two strains. Flanking a 192 bp region of the  
315 amplified sequence are low complexity regions with good partial matches for the *stx1*-forward  
316 and *stx2*-reverse primers providing a rational explanation of how a '190 bp' amplicon could be  
317 produced in a multiplex reaction. The matching sequence itself represents part of the gene  
318 encoding a DNA injecting protein of a temperate bacteriophage Gally Phage LF82\_763  
319 (Accession number OV696608), for which several *E. coli* in the database are lysogenized.



320 **Discussion**

321 **PCR performance for DEC identification is poor**

322 Diarrhoeagenic *E. coli* (DEC) are endemic in Africa where a wide range of lineages produce life-  
323 threatening infections in young children (39,40). The majority of clinical and research laboratories  
324 in West Africa are unable to detect these pathogens, therefore available evidence of their  
325 importance comes largely from short-term epidemiological surveys. Of those labs that do diagnose  
326 DEC infections, or perform surveillance, multiplex PCR is typically the technique of choice with  
327 multiple protocols published in the literature. PCR remains popular for DEC identification here  
328 and elsewhere in the world and most of the available information on DEC epidemiology comes  
329 from studies employing different PCR methodologies (8,9,42–47,10,11,17,18,29–31,41), PCR  
330 precision may vary with pathogen genome so that primers optimized for use in one part of the  
331 world may not have sufficient precision in geographies where different pathogen lineages  
332 predominate. However, there does not appear to be evidence that any have been optimized for use  
333 in West Africa where the pathogen lineage repertoire may differ from the rest of the world.

334  
335 In this study, we found that the Aranda et al (2004) (9), which performed well in its initial  
336 assessment on isolates from Brazil, has been used extensively in Nigeria, including by us, and has  
337 limited sensitivity and specificity. This is one of several multiplex PCR protocols, some of which  
338 have been well validated, but not on local African pathogenic lineages. It also includes primer  
339 pairs for EAEC, ETEC and EPEC that are used in other multiplex protocols. In our 2016-2019  
340 Ibadan, Nigeria setting, the Aranda et al (2004) (9) protocol under-estimated the prevalence of  
341 EAEC to some degree because of lowered sensitivity of the CVD432 primer for their target and to  
342 a much larger degree because too few EAEC targets are included in the multiplex. PCR

343 overestimated the prevalence of ETEC and inferred an association with disease because of non-  
344 specificity of ETEC primers and resulted in a locally predominant EAEC lineage being  
345 misclassified as ETEC. PCR led to misidentification of strains belonging to other pathotypes or  
346 commensals as Shiga toxin-producing *E. coli*, which are reportable, again because of primer non-  
347 specificity. Typical EPEC numbers were also inflated, and atypical EPEC numbers diminished  
348 due to mis-priming. Overall PCR methodology offers reasonable sensitivity or specificity, but not  
349 both, for identifying some DEC. This has implications for the use of this method in clinical  
350 diagnosis and for interpreting the results of epidemiological surveys (48). As PCR deficits are  
351 misinforming on specific lineages (such as ST720 complex) or strains carrying specific virulence  
352 gene alleles (e.g. intimin theta), the overall importance of these clades or factors is likely to be  
353 overlooked and could be significant. For example, intimin theta is among newer alleles reputed to  
354 confer enhanced invasion on atypical EPEC (49).

355  
356 That PCR is not as sensitive as WGS is not a surprise. However, when the departure in results that  
357 PCR generates is widely deviant from reality, the value of PCR testing must be questioned. We  
358 note with interest that under-performance of PCR could be attributed to different root causes for  
359 the different pathotype targets. The insensitivity diagnostic protocols that target CVD432 alone is  
360 a generalized problem with the pathotype and its current definition. The CVD432 locus, now  
361 known as *aatD* encodes part of a secretion system for the antiaggregation protein, Aap, and is  
362 present in only some EAEC. It is common among EAEC strains from some locations, for example  
363 the Chile and India strain sets on which that the “EAEC probe” Baudry et al (1990) was validated  
364 (50). However subsequent data has shown that CVD432-negative EAEC strains predominate in  
365 some geographies, including Nigeria (6,51). The insensitivity of CVD432-based PCR protocols

366 and could be resolved, at least in part by using additional targets as Czeuczulin et al and Boisen et  
367 al (52,53) have recommended. Including more primer pairs will add to the cost of detecting EAEC  
368 and, if they are multiplexed, could add risk of mis-identifications unless rigorous validation is  
369 performed.

370  
371 Sensitivity was even lower for EPEC than EAEC, although EPEC are less common overall. In  
372 this case, the number of targets cannot be faulted since EPEC strains that have other LEE genes  
373 typically have *eae*. However, we find primer mismatches in some EPEC intimin alleles that were  
374 described after the original report of intimin allelic variation by Adu-Bobie et al (1998)(28). While  
375 the divergence of these alleles from other intimins has been reported previously (54), potential  
376 mispriming at consensus *eae* priming sites has not.

377  
378 Unlike EAEC, for which the absence of a precise genetic definition requires multiple targets to be  
379 used to identify strains from this pathotype (6,32), ETEC and STEC have genetic definitions based  
380 on the presence of toxin genes that encode the virulence factors producing the disease. ETEC are  
381 vaccine priority organisms and STEC are reportable. Therefore, respective ETEC and STEC  
382 specificities in this study of 74.5% and 90.1%, respectively, are below what is needed to inform  
383 interventions and contain these pathogens.

384  
385 The major advantage of whole genome sequencing is that most DEC pathotypes can be predicted  
386 with a high degree of accuracy from sequence data (55). The obvious deterrent for routine use is  
387 cost. When isolate-based investigation of pathogens in diarrhoea stools is conducted, multiple  
388 genomes will have to be sequenced per specimen, thus WGS DEC studies are particularly

389 expensive. It is important for resource-constrained labs in endemic areas to be able to pick up key  
390 pathogens and enabling this while sacrificing some sensitivity and specificity is justifiable.  
391 Imperfect diagnostics can be useful for identification of sporadic and outbreak cases and, to some  
392 degree for epidemiologic assessments. However, it is clear that some benchmarking or validation  
393 is necessary before reliable conclusions can be drawn from DEC PCR. Some popular protocols  
394 have been validated on less than 10 strains per pathotype (18), which is far lower than the number  
395 of lineages known at the time the test was developed, a number that continues to grow. DEC  
396 identification by PCR is inexpensive and popular in resource limited settings in Africa. However,  
397 an established DEC multiplex PCR protocol is under-performing in our setting. Based on the  
398 results of this study, we would argue that the complete lack of specificity for the vaccine prioritized  
399 pathotype ETEC and misidentification of STEC and EHEC, which are reportable foodborne  
400 pathogens make this methodology, by itself, cost-inefficient. Molecular microbiology methods  
401 that are performed simply because they can be, but are unable to achieve the aims for which they  
402 were developed outside of their originally developed context are rightly currently being phased  
403 out (56). Unfortunately, there are few alternatives to end-point PCR-based DEC diagnostics that  
404 can be used broadly across West Africa, at least until WGS becomes more affordable and better  
405 disseminated. In the interim when existing multiplex PCR protocols are applied in settings  
406 different from those used to develop them, they be used alongside other methodology to verify  
407 outcome.

408

## 409 **Conclusion**

410 Pathotype lineages vary with geographic location, thus PCR sensitivity may vary based on  
411 geographic depending on the predominating pathotype lineage. Multiplex PCR is an inexpensive

412 option for identifying DEC pathotypes but cannot be relied upon to perform optimally in settings  
413 where lineages different from those on which protocols are optimized for are predominant. Where  
414 possible, epidemiological surveys for DEC should be whole-genome sequence-based. In the  
415 instances that resource constraints do not permit this additional validation steps will be required to  
416 ensure that target loci for DEC identification are appropriate and that presumptive DEC identified  
417 by PCR indeed harbor the target genes used to define them.

418

419

420

#### 421 **Acknowledgment**

422 We thank David A. Kwasi, Anderson O. Oaikhena, Taiwo Badejo, Anthony Underwood, Ayorinde  
423 Afolayan, Erkison Odih, Catherine Ladipo and A Oladipo Aboderin for technical assistance and  
424 helpful comments.

425 This work was supported by an African Research Leader's Award to INO, GD and NRT jointly  
426 funded by the UK Medical Research Council (MRC) and the UK Department for International  
427 Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2  
428 programme supported by the European Union (Award # MR/L00464X/1). INO is a Calestous  
429 Juma Science Leadership Fellow supported by the Bill and Melinda Gates Foundation (INV-  
430 036234).

431

432

433 **References**

- 434 1. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998  
435 Jan;11(1):142–201.
- 436 2. Olesen B, Scheutz F, Andersen RL, Menard M, Boisen N, Johnston B, et al.  
437 Enteroggregative *Escherichia coli* O78:H10, the Cause of an Outbreak of Urinary Tract  
438 Infection. J Clin Microbiol [Internet]. 2012 Nov;50(11):3703–11. Available from:  
439 <https://journals.asm.org/doi/10.1128/JCM.01909-12>
- 440 3. Hedberg CW, Savarino SJ, Besser JM, Paulus CJ, Thelen VM, Myers LJ, et al. An  
441 Outbreak of Foodborne Illness Caused by *Escherichia coli* O39:NM, an Agent Not Fitting  
442 into the Existing Scheme for Classifying Diarrheogenic *E. coli*. J Infect Dis [Internet].  
443 1997 Dec;176(6):1625–8. Available from: doi: 10.1086/517342
- 444 4. Aurass P, Prager R, Flieger A. EHEC/EAEC O104:H4 strain linked with the 2011 German  
445 outbreak of haemolytic uremic syndrome enters into the viable but non-culturable state in  
446 response to various stresses and resuscitates upon stress relief. Environ Microbiol  
447 [Internet]. 2011 Dec;13(12):3139–48. Available from:  
448 <https://onlinelibrary.wiley.com/doi/10.1111/j.1462-2920.2011.02604.x>
- 449 5. Dias RCB, Tanabe RHS, Vieira MA, Cergole-Novella MC, dos Santos LF, Gomes TAT,  
450 et al. Analysis of the Virulence Profile and Phenotypic Features of Typical and Atypical  
451 Enteroggregative *Escherichia coli* (EAEC) Isolated From Diarrheal Patients in Brazil.  
452 Front Cell Infect Microbiol. 2020 Apr 22;10.
- 453 6. Akinlabi OC, Nwoko EQ, Dada RA, Ekpo S, Omotuyi A, Adepoju A, et al. Epidemiology  
454 and risk factors for diarrhoeagenic *Escherichia coli* carriage among children in northern  
455 Ibadan, Nigeria. medRxiv [Internet]. 2022;2022.09.26.22280249. Available from:

- 456 <https://www.medrxiv.org/content/10.1101/2022.09.26.22280249v1%0Ahttps://www.medrxiv.org/content/10.1101/2022.09.26.22280249v1.abstract>
- 457
- 458 7. Morabito S, Karch H, Mariani-Kurkdjian P, Schmidt H, Minelli F, Bingen E, et al.
- 459 Enteroaggregative, Shiga Toxin-Producing *Escherichia coli* O111:H2 Associated with an
- 460 Outbreak of Hemolytic-Uremic Syndrome. *J Clin Microbiol* [Internet]. 1998
- 461 Mar;36(3):840–2. Available from:
- 462 <https://linkinghub.elsevier.com/retrieve/pii/S0140673605626116>
- 463 8. Antikainen J, Tarkka E, Haukka K, Siitonen A, Vaara M, Kirveskari J. New 16-plex PCR
- 464 method for rapid detection of diarrheagenic *Escherichia coli* directly from stool samples.
- 465 *Eur J Clin Microbiol Infect Dis* [Internet]. 2009 Aug 24;28(8):899–908. Available from:
- 466 <http://link.springer.com/10.1007/s10096-009-0720-x>
- 467 9. Aranda KRS, Fagundes-Neto U, Scaletsky ICA. Evaluation of Multiplex PCRs for
- 468 Diagnosis of Infection with Diarrheagenic *Escherichia coli* and *Shigella* spp. *J Clin*
- 469 *Microbiol* [Internet]. 2004 Dec;42(12):5849–53. Available from:
- 470 <https://journals.asm.org/doi/10.1128/JCM.42.12.5849-5853.2004>
- 471 10. Toma C, Lu Y, Higa N, Nakasone N, Chinen I, Baschkier A, et al. Multiplex PCR Assay
- 472 for Identification of Human Diarrheagenic *Escherichia coli*. *J Clin Microbiol* [Internet].
- 473 2003 Jun;41(6):2669–71. Available from:
- 474 <https://journals.asm.org/doi/10.1128/JCM.41.6.2669-2671.2003>
- 475 11. López-Saucedo C, Cerna JF, Villegas-Sepulveda N, Thompson R, Velazquez FR, Torres
- 476 J, et al. Single Multiplex Polymerase Chain Reaction To Detect Diverse Loci Associated
- 477 with Diarrheagenic *Escherichia coli*. *Emerg Infect Dis* [Internet]. 2003 Jan;9(1):127–31.
- 478 Available from: [http://wwwnc.cdc.gov/eid/article/9/1/01-0507\\_article.htm](http://wwwnc.cdc.gov/eid/article/9/1/01-0507_article.htm)

- 479 12. Scaletsky ICA, Fabbriotti SH, Aranda KR, Morais MB, Fagundes-Neto U. Comparison  
480 of DNA Hybridization and PCR Assays for Detection of Putative Pathogenic  
481 Enteroadherent *Escherichia coli*. *J Clin Microbiol* [Internet]. 2002 Apr;40(4):1254–8.  
482 Available from: <https://journals.asm.org/doi/10.1128/JCM.40.4.1254-1258.2002>
- 483 13. Snelling AM, Macfarlane-Smith LR, Fletcher JN, Okeke IN. The commonly-used DNA  
484 probe for diffusely-adherent *Escherichia coli* cross-reacts with a subset of  
485 enteroaggregative *E. coli*. *BMC Microbiol* [Internet]. 2009 Dec 21;9(1):269. Available  
486 from: <https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-9-269>
- 487 14. Odetoyin BW, Hofmann J, Aboderin AO, Okeke IN. Diarrhoeagenic *Escherichia coli* in  
488 mother-child Pairs in Ile-Ife, South Western Nigeria. *BMC Infect Dis* [Internet]. 2015 Dec  
489 25;16(1):28. Available from: <http://www.biomedcentral.com/1471-2334/16/28>
- 490 15. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective  
491 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*  
492 [Internet]. 2015 Jan;32(1):268–74. Available from: [https://academic.oup.com/mbe/article-](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300)  
493 [lookup/doi/10.1093/molbev/msu300](https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msu300)
- 494 16. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact:  
495 visualizing and sharing data for genomic epidemiology and phylogeography. *Microb*  
496 *Genomics* [Internet]. 2016 Nov 30;2(11). Available from:  
497 <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000093>
- 498 17. Tobias J, Vutukuru S-R. Simple and rapid multiplex PCR for identification of the main  
499 human diarrheagenic *Escherichia coli*. *Microbiol Res* [Internet]. 2012 Oct;167(9):564–70.  
500 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S094450131100111X>
- 501 18. Vidal M, Kruger E, Durán C, Lagos R, Levine M, Prado V, et al. Single Multiplex PCR



- 502 Assay To Identify Simultaneously the Six Categories of Diarrheogenic Escherichia coli  
503 Associated with Enteric Infections. J Clin Microbiol [Internet]. 2005 Oct;43(10):5362–5.  
504 Available from: <https://journals.asm.org/doi/10.1128/JCM.43.10.5362-5365.2005>
- 505 19. Müller D, Greune L, Heusipp G, Karch H, Fruth A, Tschäpe H, et al. Identification of  
506 Unconventional Intestinal Pathogenic Escherichia coli Isolates Expressing Intermediate  
507 Virulence Factor Profiles by Using a Novel Single-Step Multiplex PCR. Appl Environ  
508 Microbiol [Internet]. 2007 May 15;73(10):3380–90. Available from:  
509 <https://journals.asm.org/doi/10.1128/AEM.02855-06>
- 510 20. El Metwally HAR, Ibrahim HAH, El-Athamna MN, Amer MA. Multiplex PCR for  
511 detection of diarrheogenic Escherichia coli in Egyptian children. J Med Sci.  
512 2007;7(2):255–62.
- 513 21. Hegde A, Ballal M, Shenoy S. Detection of diarrheogenic Escherichia coli by multiplex  
514 PCR. Indian J Med Microbiol [Internet]. 2012 Jul;30(3):279–84. Available from:  
515 <https://linkinghub.elsevier.com/retrieve/pii/S0255085721009063>
- 516 22. Fialho OB, de Souza EM, de Borba Dallagassa C, de Oliveira Pedrosa F, Klassen G, Irino  
517 K, et al. Detection of Diarrheogenic Escherichia coli Using a Two-System Multiplex-PCR  
518 Protocol. J Clin Lab Anal [Internet]. 2013 Mar;27(2):155–61. Available from:  
519 <https://onlinelibrary.wiley.com/doi/10.1002/jcla.21578>
- 520 23. Helmy OM, Ragab YM, Hussein MMM. Multiplex polymerase chain reaction ( PCR ) for  
521 the detection of diarrheogenic Escherichia coli and Shigella directly from stool. African J  
522 Microbiol Res. 2013;7(34):4368–72.
- 523 24. Al-Talib H, Latif B, Mohd-Zain Z. Pentaplex PCR assay for detection of hemorrhagic  
524 bacteria from stool samples. J Clin Microbiol. 2014;52(9):3244–9.

- 525 25. Oh K-H, Kim DW, Jung S-M, Cho S-H. Molecular Characterization of Enterotoxigenic  
526 Escherichia coli Strains Isolated from Diarrheal Patients in Korea during 2003–2011.  
527 Vadivelu J, editor. PLoS One [Internet]. 2014 May 19;9(5):e96896. Available from:  
528 <https://dx.plos.org/10.1371/journal.pone.0096896>
- 529 26. Onori M, Coltella L, Mancinelli L, Argentieri M, Menichella D, Villani A, et al.  
530 Evaluation of a multiplex PCR assay for simultaneous detection of bacterial and viral  
531 enteropathogens in stool samples of paediatric patients. Diagn Microbiol Infect Dis  
532 [Internet]. 2014 Jun;79(2):149–54. Available from:  
533 <https://linkinghub.elsevier.com/retrieve/pii/S0732889314000674>
- 534 27. Naziri Z, Derakhshandeh A, Firouzi R, Motamedifar M, Shojaee Tabrizi A. DNA  
535 fingerprinting approaches to trace Escherichia coli sharing between dogs and owners. J  
536 Appl Microbiol [Internet]. 2016 Feb;120(2):460–8. Available from:  
537 <https://onlinelibrary.wiley.com/doi/10.1111/jam.13003>
- 538 28. Sjöling Å, Sadeghipoorjahromi L, Novak D, Tobias J. Detection of major diarrheagenic  
539 bacterial pathogens by multiplex PCR panels. Microbiol Res [Internet]. 2015 Mar;172:34–  
540 40. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0944501314001517>
- 541 29. Zhang J, Xu Y, Ling X, Zhou Y, Lin Z, Huang Z, et al. Identification of diarrheagenic  
542 Escherichia coli by a new multiplex PCR assay and capillary electrophoresis. Mol Cell  
543 Probes [Internet]. 2020 Feb;49:101477. Available from:  
544 <https://linkinghub.elsevier.com/retrieve/pii/S089085081930324X>
- 545 30. Hegde A, Ballal M, Shenoy S. Detection of diarrheagenic Escherichia coli by multiplex  
546 PCR. Indian J Med Microbiol [Internet]. 2012 Jul;30(3):279–84. Available from:  
547 <https://linkinghub.elsevier.com/retrieve/pii/S0255085721009063>

- 548 31. Opintan JA, Bishar RA, Newman MJ, Okeke IN. Carriage of diarrhoeagenic *Escherichia*  
549 *coli* by older children and adults in Accra, Ghana. *Trans R Soc Trop Med Hyg*. 2010  
550 Jul;104(7):504–6.
- 551 32. Boisen N, Østerlund MT, Joensen KG, Santiago AE, Mandomando I, Cravioto A, et al.  
552 Redefining enteroaggregative *Escherichia coli* (EAEC): Genomic characterization of  
553 epidemiological EAEC strains. *PLoS Negl Trop Dis*. 2020 Sep 8;14(9):e0008613.
- 554 33. Nazari Z, Amini K, Mokhtari A. Original article Detection of diarrhoeagenic *Escherichia*  
555 *coli* virulence genes by multiplex-PCR method and their antibiotic susceptibility profile.  
556 2015;4:40–8.
- 557 34. Puño-Sarmiento J, Gazal L, Medeiros L, Nishio E, Kobayashi R, Nakazato G.  
558 Identification of Diarrheagenic *Escherichia coli* Strains from Avian Organic Fertilizers.  
559 *Int J Environ Res Public Health* [Internet]. 2014 Aug 28;11(9):8924–39. Available from:  
560 <http://www.mdpi.com/1660-4601/11/9/8924>
- 561 35. Dias RCB, dos Santos BC, dos Santos LF, Vieira MA, Yamatogi RS, Mondelli AL, et al.  
562 Diarrheagenic *Escherichia coli* pathotypes investigation revealed atypical  
563 enteropathogenic *E. coli* as putative emerging diarrheal agents in children living in  
564 Botucatu, São Paulo State, Brazil. *APMIS* [Internet]. 2016 Apr;124(4):299–308. Available  
565 from: <https://onlinelibrary.wiley.com/doi/10.1111/apm.12501>
- 566 36. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in  
567 *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* [Internet]. 2006  
568 Jun;60(5):1136–51. Available from: [https://onlinelibrary.wiley.com/doi/10.1111/j.1365-](https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2006.05172.x)  
569 [2958.2006.05172.x](https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2958.2006.05172.x)
- 570 37. Prah I, Ayibieke A, Nguyen TTH, Iguchi A, Mahazu S, Sato W, et al. Virulence profiles

- 571 of diarrheagenic *Escherichia coli* isolated from the western region of Ghana. *Jpn J Infect*  
572 *Dis.* 2021;74(2):115–21.
- 573 38. Greig DR, Jenkins C, Gharbia S, Dallman TJ. Comparison of single-nucleotide variants  
574 identified by Illumina and Oxford Nanopore technologies in the context of a potential  
575 outbreak of Shiga toxin–producing *Escherichia coli*. *Gigascience* [Internet]. 2019 Aug  
576 1;8(8). Available from:  
577 <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz104/5552683>
- 578 39. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, et al.  
579 Burden and aetiology of diarrhoeal disease in infants and young children in developing  
580 countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control  
581 study. *Lancet.* 2013 Jul;382(9888):209–22.
- 582 40. Rogawski ET, Guerrant RL, Havt A, Lima IFN, Medeiros PHQS, Seidman JC, et al.  
583 Epidemiology of enteroaggregative *Escherichia coli* infections and associated outcomes in  
584 the MAL-ED birth cohort. *PLoS Negl Trop Dis.* 2017 Jul 24;11(7):e0005798.
- 585 41. Mokomane M, Kasvosve I, Melo E de, Pernica JM, Goldfarb DM. The global problem of  
586 childhood diarrhoeal diseases: emerging strategies in prevention and management. *Ther*  
587 *Adv Infect Dis.* 2018 Jan 7;5(1):29–43.
- 588 42. Mosquito S, Pons MJ, Riveros M, Ruiz J, Ochoa TJ. Diarrheagenic *Escherichia coli*  
589 Phylogroups Are Associated with Antibiotic Resistance and Duration of Diarrheal  
590 Episode. *Sci World J.* 2015;2015:1–6.
- 591 43. Jerse AE, Kaper JB. The *eae* gene of enteropathogenic *Escherichia coli* encodes a 94-  
592 kilodalton membrane protein, the expression of which is influenced by the EAF plasmid.  
593 *Infect Immun* [Internet]. 1991 Dec;59(12):4302–9. Available from:

- 594 <https://journals.asm.org/doi/10.1128/iai.59.12.4302-4309.1991>
- 595 44. Onanuga A, Igbeneghu O, Lamikanra A. A study of the prevalence of diarrhoeagenic  
596 *Escherichia coli* in children from Gwagwalada, Federal Capital Territory, Nigeria. *Pan Afr*  
597 *Med J*. 2014;17.
- 598 45. Ali MMM, Ahmed SF, Klena JD, Mohamed ZK, Moussa TA, Ghenghesh KS.  
599 Enteraggregative *Escherichia coli* in diarrheic children in Egypt: molecular  
600 characterization and antimicrobial susceptibility. *J Infect Dev Ctries*. 2014 May  
601 14;8(05):589–96.
- 602 46. Sullivan, Bolton D. J, G. Duffy, C. Baylis, R. Tozzoli YW and SL. Methods for Detection  
603 and Molecular Characterisation of Pathogenic *Escherichia coli*. *Pathog Escherichia coli*  
604 *Netw*. 2006;2–29.
- 605 47. Okeke IN, Ojo O, Lamikanra A, Kaper JB. Etiology of Acute Diarrhea in Adults in  
606 Southwestern Nigeria. *J Clin Microbiol*. 2003 Oct;41(10):4525–30.
- 607 48. Jesser KJ, Levy K. Updates on defining and detecting diarrheagenic *Escherichia coli*  
608 pathotypes. *Curr Opin Infect Dis [Internet]*. 2020 Oct;33(5):372–80. Available from:  
609 <https://journals.lww.com/10.1097/QCO.0000000000000665>
- 610 49. Yamamoto D, Hernandes RT, Blanco M, Greune L, Schmidt MA, Carneiro SM, et al.  
611 Invasiveness as a putative additional virulence mechanism of some atypical  
612 Enteropathogenic *Escherichia coli* strains with different uncommon intimin types. *BMC*  
613 *Microbiol [Internet]*. 2009 Dec 21;9(1):146. Available from:  
614 <https://bmcmicrobiol.biomedcentral.com/articles/10.1186/1471-2180-9-146>
- 615 50. Baudry B, Savarino SJ, Vial P, Kaper JB, Levine MM. A Sensitive and Specific DNA  
616 Probe to Identify Enteraggregative *Escherichia coli*, a Recently Discovered Diarrheal

- 617 Pathogen. J Infect Dis [Internet]. 1990 Jun 1;161(6):1249–51. Available from:  
618 <https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/161.6.1249>
- 619 51. Okeke IN, Lamikanra A, Czeczulin J, Dubovsky F, Kaper JB, Nataro JP. Heterogeneous  
620 Virulence of Enteroaggregative *Escherichia coli* Strains Isolated from Children in  
621 Southwest Nigeria. J Infect Dis [Internet]. 2000 Jan;181(1):252–60. Available from:  
622 <https://academic.oup.com/jid/article-lookup/doi/10.1086/315204>
- 623 52. Czeczulin JR, Whittam TS, Henderson IR, Navarro-Garcia F, Nataro JP. Phylogenetic  
624 Analysis of Enteroaggregative and Diffusely Adherent *Escherichia coli*. Infect Immun.  
625 1999 Jun;67(6):2692–9.
- 626 53. Boisen N, Scheutz F, Rasko DA, Redman JC, Persson S, Simon J, et al. Genomic  
627 Characterization of Enteroaggregative *Escherichia coli* From Children in Mali. J Infect  
628 Dis. 2012 Feb 1;205(3):431–44.
- 629 54. Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, et al. Clinical Significance  
630 of *Escherichia albertii*. Emerg Infect Dis [Internet]. 2012 Mar;18(3):488–92. Available  
631 from: [http://wwwnc.cdc.gov/eid/article/18/3/11-1401\\_article.htm](http://wwwnc.cdc.gov/eid/article/18/3/11-1401_article.htm)
- 632 55. Robins-Browne RM, Holt KE, Ingle DJ, Hocking DM, Yang J, Tauschek M. Are  
633 *Escherichia coli* Pathotypes Still Relevant in the Era of Whole-Genome Sequencing?  
634 Front Cell Infect Microbiol [Internet]. 2016 Nov 18;6. Available from:  
635 <http://journal.frontiersin.org/article/10.3389/fcimb.2016.00141/full>
- 636 56. Achtman M. A surfeit of YATMs? [3]. J Clin Microbiol. 1996;34(7):1870.

638

639 **Figure Legends**

640 Figure 1: A selection of multiplex PCR gels (a) Multiplex PCR1 with lane L, 1-kb size ladder,  
641 Lane 2, EAEC 042 reference control, Lane N, negative control, lane 3-4 *E. coli* strains  
642 INOLLD97B and INOLLH029A positive for *eae* gene, lane 6, 11, 12, positive isolated *E. coli*  
643 strains with *bfp* gene, Lane 7, isolated *E. coli* strains with the *bfp* and *eae* genes, lane 10, EPEC  
644 E2348/69 reference control with the *bfp* and *eae* genes, lane 13 and 14 *E. coli* strains  
645 INOLWH056C and INOLLH297C positive for CVD432 gene (b) Multiplex PCR3 with lane L, 1-  
646 kb size ladder, Lane N, negative control, Lane 1, EIEC E137 reference control with *ipaH* gene,  
647 lane 2 ETEC H10407 reference control with the *elt* and *est* genes, lane 3-4 isolates INOMNH199C  
648 and INOLLD018D *E. coli* strains positive for *est* gene. (c) Multiplex PCR4 with EHEC EDL933  
649 reference controls (Lane C). Lane 1-5, strains INOMND61D, INOLLH045A, INOLLH030F,  
650 INOMNH077E and INOLKH217A with *stx2* genes, Lane L, 1-kb size ladder.

651

652 Figure 2: Maximum likelihood SNP tree of genomes used in this evaluation showing sequence  
653 type, results for *stx1*, *stx2*, *stx1+2* PCR positivity; *elt*, *est* and *elt+est* PCR and WGS positivity,  
654 and CVD432 PCR and WGS positivity. The figure can be engaged interactively at  
655 <https://microreact.org/project/pWbLpY3qhV6Npa9q8Rpgrk-wgs-n-pcr>. While there was some  
656 concordance for CVD432, there was little for the STEC and ETEC toxins. Nodes for the ST720  
657 and ST69 complexes in which *elt* false positives are over-represented are indicated with arrows.  
658 (b) False and true positives, and negatives, for *elt* overall and for ST720 and ST69 complexes.

659

660 Figure 3: Amplicon sequence with matches >6 contiguous base pairs on any *stx* primer  
661 highlighted. (b) Corresponding highlighted regions in *stx* primers

Figure 1

medRxiv preprint doi: <https://doi.org/10.1101/2023.01.06.23284276>; this version posted January 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

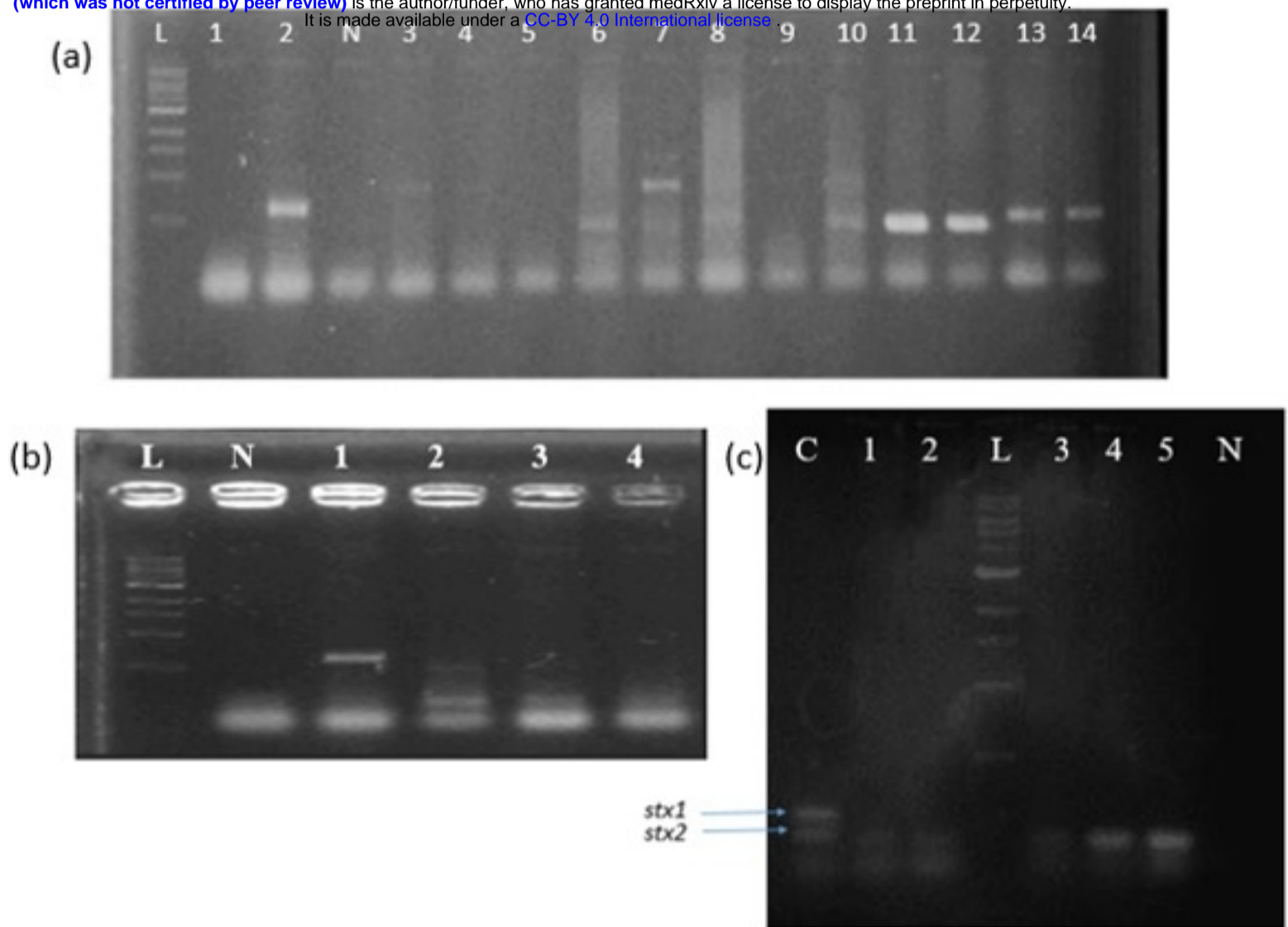
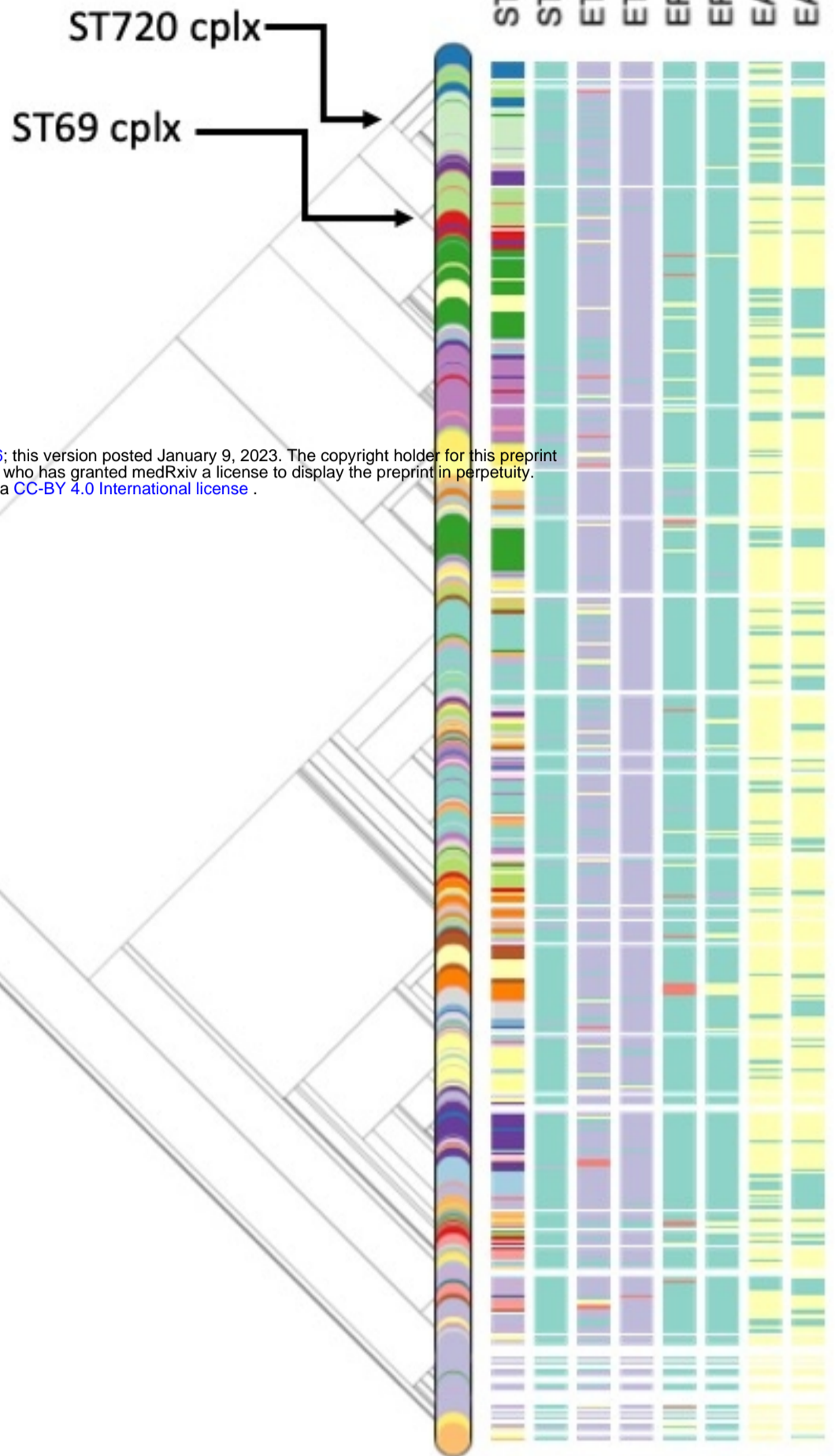


Figure 1



Figure 2

A



medRxiv preprint doi: <https://doi.org/10.1101/2023.01.06.23284276>; this version posted January 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Colours by STEC PCR

- N
- stx1n2
- stx2

Colours by ETEC PCR

- LT
- LT,ST
- N
- ST

Colours by ETEC WGS

- LT
- LT,ST
- N
- ST

Colours by EPEC PCR

- bfp
- bfp eae
- eae

Colours by EPEC WGS

- N
- eae
- eae bfp
- N

Colours by EAEC PCR

- CVD432
- N

Colours by EAEC WGS

- CVD432
- N

B

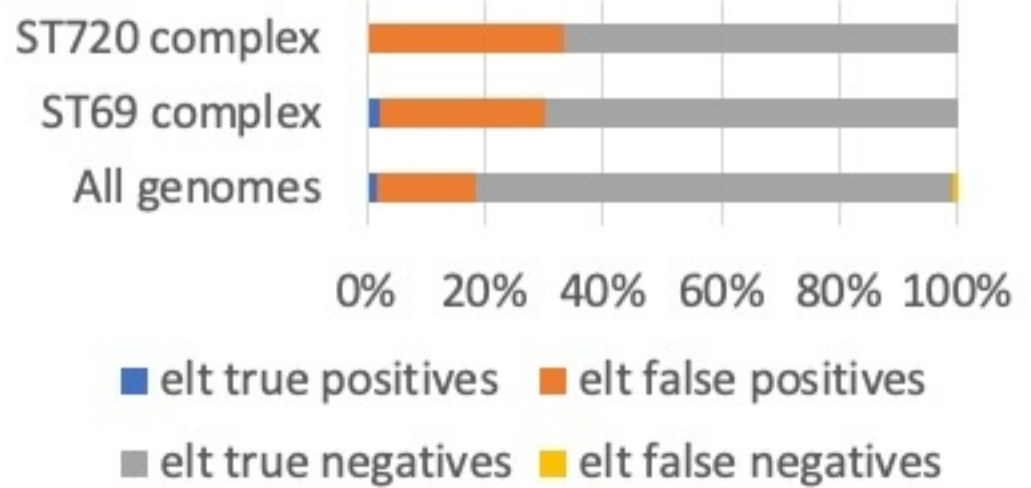


Figure 2

